

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

CORR. TO US 6,654,722

(11) 特許出願公開番号

特開2002-57724

(P2002-57724A)

(43) 公開日 平成14年2月22日 (2002.2.22)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テーマコード* (参考)
H 0 4 L 12/66		H 0 4 L 12/66	D 5 K 0 3 0
H 0 4 M 11/00	3 0 2	H 0 4 M 11/00	3 0 2 5 K 1 0 1

審査請求 有 請求項の数21 O L (全 12 頁)

(21) 出願番号 特願2001-184053(P2001-184053)

(22) 出願日 平成13年6月18日 (2001.6.18)

(31) 優先権主張番号 09/596769

(32) 優先日 平成12年6月19日 (2000.6.19)

(33) 優先権主張国 米国 (US)

(71) 出願人 390009531

インターナショナル・ビジネス・マシーンズ・コーポレーション

INTERNATIONAL BUSINESS MACHINES CORPORATION

アメリカ合衆国10504、ニューヨーク州

アーモンク (番地なし)

(74) 代理人 100086243

弁理士 坂口 博 (外2名)

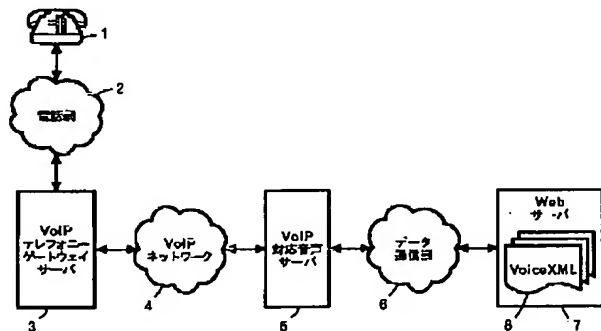
最終頁に続く

(54) 【発明の名称】 VoIPプロトコル・ベースの音声システム

(57) 【要約】

【課題】 音声アプリケーションの設計および実装が、IPテレフォニー・システムの設計および実装から分離した状態を維持するVoIPベースの音声システムを提供すること。

【解決手段】 VoIP対応音声サーバは、VoIP通信パスを介してVoIPテレフォニー・ゲートウェイ・サーバと通信するように構成され得る音声アプリケーションを含む。また、VoIP対応音声サーバは、VoIPテレフォニー・ゲートウェイ・サーバへのVoIP対応呼び制御インタフェースを含み、VoIP対応呼び制御インタフェースはVoIP通信パスを構築する。動作時において、音声アプリケーションは、VoIP対応パケットを、VoIP通信パスを介して受信する。その結果、デジタル化音声データは、VoIP対応パケットから再構成することができ、デジタル化音声データは音声-テキスト変換される。さらに、テキストは、デジタル化音声データに合成され、デジタル化音声データはVoIP対応パケットにカプセル化され、VoIP通信パスを介して、テレフォニー・ゲートウェイ・サーバに送信される。



【特許請求の範囲】

【請求項1】 ボイス・オーバーIP (VoIP) ベースの音声システムであって、

VoIPテレフォニー・ゲートウェイ・サーバと、それぞれが音声アプリケーションを含む少なくとも1つの音声サーバと、

前記VoIPテレフォニー・ゲートウェイ・サーバと前記音声サーバとの間のVoIP対応制御インタフェースと、

前記VoIPテレフォニー・ゲートウェイ・サーバと前記少なくとも1つの音声サーバ中の前記音声アプリケーションとの間のVoIP通信パスとを含み、

前記VoIPテレフォニー・ゲートウェイ・サーバと前記音声アプリケーションとは、前記VoIP対応制御インタフェースを介して前記VoIP通信パスを構築し、

前記VoIPテレフォニー・ゲートウェイ・サーバは、テレフォニー・インタフェースから音声信号を受信し、前記音声信号をデジタル化音声データにデジタル化し、前記デジタル化音声データをVoIP対応パケットに圧縮し、前記VoIP対応パケットを、前記少なくとも1つの音声サーバ中の前記音声アプリケーションへ、前記VoIP通信パスを介して、前記VoIPプロトコルを使用して送信し、

前記少なくとも1つの音声サーバ中の前記音声アプリケーションは、前記VoIP対応パケットを受信し、前記デジタル化音声データを前記VoIP対応パケットから再構成し、前記デジタル化音声データを音声テキスト変換し、

前記少なくとも1つの音声サーバ中の前記音声アプリケーションは、テキストをデジタル化音声データに音声合成し、前記デジタル化音声データをVoIP対応パケットにカプセル化し、前記VoIP対応パケットを、前記VoIP通信パスを介して、前記VoIPテレフォニー・ゲートウェイ・サーバへ送信し、

前記VoIPテレフォニー・ゲートウェイ・サーバは、前記VoIP対応パケットを受信し、前記デジタル化音声データを、前記VoIP対応パケットから再構成し、前記デジタル化音声データを、前記テレフォニー・インタフェースを介して送信する、VoIPベースの音声システム。

【請求項2】 前記VoIPテレフォニー・ゲートウェイ・サーバは、

テレフォニー・インタフェースと、VoIPゲートキーパーとを含み、

前記VoIPゲートキーパーは、前記テレフォニー・インタフェースを介して音声呼を受信し、

前記VoIPゲートキーパーは、前記少なくとも1つの音声サーバから1つの音声サーバを選択し、

前記VoIPゲートキーパーは、選択音声サーバ中の前

記VoIP対応音声アプリケーションに対して、前記音声呼が受信された旨をアラートする、

請求項1記載のVoIPベースの音声システム。

【請求項3】 前記少なくとも1つの音声サーバは、音声認識エンジンと、

テキスト音声変換エンジンと、

前記VoIPテレフォニー・ゲートウェイ・サーバを介して音声呼接続を構築する制御インタフェースと、音声データ・パスとを含み、

前記音声データ・パスは、構築された音声呼接続を前記音声認識エンジンへストリーミングし、

前記音声データ・パスは、前記テキスト音声変換エンジンから前記構築された音声呼接続を介して、音声データをストリーミングする、

請求項1記載のVoIPベースの音声システム。

【請求項4】 前記音声アプリケーションは、音声ブラウザであり、前記音声ブラウザは、前記VoIP通信パスを介して受信された音声コマンドにตอบสนองして、Webコンテンツを検索し、前記音声ブラウザは、検索されたWebコンテンツを音声合成化して音声データを生成し、前記音声ブラウザは、前記音声データを、前記VoIP通信パスを介して、前記VoIPテレフォニー・ゲートウェイ・サーバへ送信する、請求項1記載のVoIPベースの音声システム。

【請求項5】 前記Webコンテンツは、VoiceXMLドキュメントである、請求項4記載のVoIPベースの音声システム。

【請求項6】 前記少なくとも1つの音声サーバは、音声認識エンジンと、

テキスト音声変換エンジンと、

J SAPI 音声インタフェースと、

J TAPI テレフォニー・インタフェースと、

J MFメディア・インタフェースとを含み、

前記J TAPI テレフォニー・インタフェースは、デジタル音声データを、前記VoIPテレフォニー・ゲートウェイ・サーバと前記音声アプリケーションとの間でトランスポートするための音声呼接続を構築し、

前記J MFメディア・インタフェースは、前記デジタル音声データを、前記音声アプリケーションと前記音声呼接続との間でトランスポートするためのデータ・パスを構築し、

前記J SAPI 音声インタフェースは、前記デジタル化音声データを、前記音声アプリケーションから前記音声認識エンジンに伝達し、

前記J SAPI 音声インタフェースは、音声合成化音声データを、前記テキスト音声変換エンジンから、前記音声アプリケーションへ伝達する、

請求項1記載のVoIPベースの音声システム。

【請求項7】 VoIP通信パスを介してVoIPテレフォニー・ゲートウェイ・サーバと通信するように構成

された音声アプリケーションと、  
前記V o I Pテレフォニー・ゲートウェイ・サーバへの  
V o I P対応呼制御インタフェースとを含み、  
前記V o I P対応呼制御インタフェースは前記V o I P  
通信パスを構築し、  
前記音声アプリケーションは、V o I P対応パケット  
を、前記V o I Pテレフォニー・ゲートウェイ・サーバ  
から、前記V o I P通信パスを介して受信し、デジタル  
化音声データを、前記V o I P対応パケットから再構  
成し、前記デジタル化音声データを音声-テキスト変  
換する、

V o I P対応音声サーバ。

【請求項8】 前記音声アプリケーションは、テキスト  
をデジタル化音声データに合成し、前記デジタル化  
音声データを、V o I P対応パケットにカプセル化し、  
前記V o I P対応パケットを、前記V o I P通信パスを  
介して、前記テレフォニー・ゲートウェイ・サーバへ送  
信する、請求項7記載のV o I P対応音声サーバ。

【請求項9】 音声認識エンジンと、  
テキスト-音声変換エンジンと、  
音声データ・パスとをさらに含み、  
前記音声データ・パスは、音声データを前記構築された  
音声呼接続を介して前記音声認識エンジンにストリー  
ミングし、  
前記音声データ・パスは、前記構築された音声呼接続を  
介して前記テキスト-音声変換エンジンから音声データ  
をストリーミングする、請求項7記載のV o I P対応音  
声サーバ。

【請求項10】 前記音声アプリケーションは音声ブラ  
ウザであり、前記音声ブラウザは、前記V o I P通信パ  
スを介して受信された音声コマンドに回答してWe bコ  
ンテンツを検索し、前記音声ブラウザは、検索されたW  
e bコンテンツを、音声データに音声合成化し、前記音  
声ブラウザは、前記音声データを、前記V o I P通信パ  
スを介して、前記V o I Pテレフォニー・ゲートウェイ  
・サーバへ送信する、請求項7記載のV o I P対応音声  
サーバ。

【請求項11】 前記We bコンテンツはV o i c e X  
M Lドキュメントである、請求項10記載のV o I P  
対応音声サーバ。

【請求項12】 前記V o I P対応呼制御インタフェ  
ースは、前記V o I Pテレフォニー・ゲートウェイ・サー  
バと前記音声アプリケーションとの間でデジタル音声  
データをトランスポートするために音声呼接続を構築す  
るJ T A P Iテレフォニー・インタフェースである、請  
求項7記載のV o I P対応音声サーバ。

【請求項13】 音声認識エンジンと、  
テキスト-音声エンジンと、  
J S A P I音声インタフェースと、  
J M Fメディア・インタフェースとを含み、

前記J M Fメディア・インタフェースは、前記ディジ  
タル音声データを、前記音声アプリケーションと前記音  
声呼接続との間でトランスポートするためのデータ・パス  
を構築し、

前記J S A P I音声インタフェースは、前記ディジ  
タル音声データを、前記音声アプリケーションから前記音  
声認識エンジンに伝達し、

前記J S A P I音声インタフェースは、音声合成化音  
声データを、前記テキスト-音声変換エンジンから、前記  
音声アプリケーションへ伝達する、

請求項12記載のV o I Pベースの対応音声サーバ。

【請求項14】 音声アプリケーションをV o I Pネッ  
トワーク中のV o I Pテレフォニー・ゲートウェイ・サ  
ーバに結合する方法であって、

V o I Pテレフォニー・ゲートウェイ・サーバとのV o  
I P通信パスを構築し、音声アプリケーションを、構築  
されたV o I P通信パスを介してV o I Pテレフォニー  
・ゲートウェイ・サーバと通信するように構成するステ  
ップと、

前記構築されたV o I P通信パスを介して、V o I Pテ  
レフォニー・ゲートウェイ・サーバからV o I P対応パ  
ケットを受信するステップと、

デジタル化音声データを前記V o I P対応パケットか  
ら再構成するステップと、

前記デジタル化音声データを音声-テキスト変換する  
ステップと、  
を含む方法。

【請求項15】 テキストをデジタル化音声データに  
合成するステップと、

前記デジタル化音声データをV o I P対応パケットに  
カプセル化するステップと、

前記V o I P対応パケットを、前記V o I P通信パスを  
介して、V o I Pテレフォニー・ゲートウェイ・サーバ  
へ送信するステップと、

をさらに含む請求項14記載の方法。

【請求項16】 前記V o I P通信パスを介して受信さ  
れた音声認識音声コマンドに回答してWe bコンテンツ  
を検索するステップと、

前記検索We bコンテンツを音声データに合成するステ  
ップと、

前記音声データを、前記V o I P通信パスを介して、V  
o I Pテレフォニー・ゲートウェイ・サーバへ送信する  
ステップと、

をさらに含む請求項14記載の方法。

【請求項17】 前記We bコンテンツは、V o i c e  
X M Lドキュメントである、請求項16記載の方法。

【請求項18】 音声アプリケーションをV o I Pネッ  
トワーク中のV o I Pテレフォニー・ゲートウェイ・サ  
ーバに結合するためのコンピュータ・プログラムを格納  
した機械可読記憶媒体であって、前記コンピュータ・プ

ログラムは、

V o I Pテレフォニー・ゲートウェイ・サーバとのV o I P通信パスを構築し、音声アプリケーションを、構築されたV o I P通信パスを介してV o I Pテレフォニー・ゲートウェイ・サーバと通信するように構成するステップと、

V o I P対応パケットを、V o I Pテレフォニー・ゲートウェイ・サーバから、前記構築されたV o I P通信パスを介して受信するステップと、

前記V o I P対応パケットからデジタル化音声データを再構成するステップと、

前記デジタル化音声データを音声-テキスト変換するステップと、

を機械に実行させる、機械可読記憶媒体。

【請求項19】 テキストを合成してデジタル化音声データを生成するステップと、

前記デジタル化音声データをV o I P対応パケットにカプセル化するステップと、

前記V o I P対応パケットを、前記V o I P通信パスを介して、V o I Pテレフォニー・ゲートウェイ・サーバへ送信するステップと、

を機械にさらに実行させる、請求項18記載の機械可読記憶媒体。

【請求項20】 前記V o I P通信パスを介して受信され音声認識された音声コマンドにตอบสนองして、Webコンテンツを検索するステップと、

前記検索されたWebコンテンツを合成して音声データを生成するステップと、

前記音声データを、前記V o I P通信パスを介して、V o I Pテレフォニー・ゲートウェイ・サーバへ送信するステップと、

をさらに機械に実行させる、請求項18記載の機械可読記憶媒体。

【請求項21】 前記WebコンテンツはVoiceXMLドキュメントである、請求項20記載の機械可読記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、音声認識の分野に関し、より詳細に述べると、ボイス・オーバーIP (V o I P) プロトコル・ネットワークで使用される音声アプリケーションに関する。

【0002】

【従来の技術】LANテレフォニーは、「パケット交換データ網により提供されるテレフォニーおよびデータ・サービスの統合」を意味するものであり、個人間通信を向上させ、関連する費用を削減させる技術である。LANテレフォニーによって、たとえば、自動呼分配、対話型音声応答、音声ロギングなどの多くのアプリケーションの使用がさらに柔軟になり、費用対効果も向上させる

ことが可能となる。これは、現在の音声/データ統合パラダイムにより提供される比較的制限された統合とは逆のもので、音声トラフィックがデータ・トラフィックと個別な状態が維持され、回線交換リンク上で運ばれるコンピュータ・テレフォニー統合である。データと音声の統合に対する古いパラダイムは、データ通信用の回線交換テレフォニー構造を使用してきたが、データ・トラフィックで使用可能な帯域が比較的低いこと、データ・トラフィックの「バースト性」に起因して回線交換データ通信が低いこと、さらに音声/データ統合の可能性が制限されることなどの明らかな欠点によって、IPデータ・サーバが私有PBXまたは音声回線交換とバンドルされて、回線交換網とパケット交換網との間の統合の程度が低下し、音声は回線交換網によって運ばれるという現在のトポロジーができた。

【0003】LANテレフォニーの最も一般的な使用の1つは、IPテレフォニーと呼ばれる企業インターネット/イントラネット環境において、である。インターネットを使用した音声通信 (ボイス・オーバーIP、以下V o I Pという) プロトコルは、音声トラフィックをIP網上で伝送することが可能なプロトコルである。

【0004】V o I Pネットワークにおいて、たとえばPSTNまたはマイクロフォンなどのアナログ音声源から受信したアナログ音声信号は、デジタル化され、圧縮され、さらにIPパケットに変換されて、IPネットワークを介して送信される。H. 232、セッション開始プロトコル (SIP)、主ゲートウェイ制御プロトコル (MGCP) を含む、周知のいくつかのプロトコルは、V o I Pプロトコル仕様をインプリメントする。

【0005】IPテレフォニーに対する一般的なアプリケーションは、ボイスメール (vメール) および電子メール (eメール) である。別のアプリケーションは、金融機関または緊急応答機関による音声ロギングを含む。さらに、自動呼分配 (ACD) は、ACDサーバが着信電話呼のバリュー・ベースのキューイングを行うように適応できる。また、対話型音声応答システムは、応答がワークフロー・コンポーネントとしてサーバ中にプログラミングされるIPテレフォニーを導入することが可能である。さらに、音声認識アプリケーションおよび音声合成アプリケーション (音声アプリケーション) は、IPテレフォニーの使用において、遅れている。

【0006】特に、音声アプリケーションは、従来のデータ通信に関連する待ち時間を許容できない実時間音声信号に対して実行される。そのため、音声アプリケーションがIPテレフォニーのトポロジー中に導入された場合、この音声アプリケーションは、IPテレフォニー・サーバと密に統合され、ネットワークに基づく待ち時間からの悪影響を防ぐ。したがって、このようなIPテレフォニー対応音声アプリケーションの設計および開発は、IPテレフォニー・サーバの私有性に密につながって

いる。

【0007】音声アプリケーションとIPテレフォニー・サーバとの間の密なつながりは、音声アプリケーションの設計と拡張性との両方を非常に制限する。特に、現在のパラダイムにおいて、音声アプリケーション設計は、パケット化された音声データを音声認識システムへ、および音声合成システムからトランスポートするための選択プロトコルに直接関係した機能を取り入れなければならない。IPテレフォニー・サーバと音声アプリケーションとの間の密なつながりのため、優れた音声トランスポート・プロトコルを開発するには、音声アプリケーションの再設計が不可欠である。したがって、音声アプリケーションの設計および実装が、IPテレフォニー・システムの設計および開発から分離した状態を維持するVoIPベースの音声システムが必要である。

【0008】

【発明が解決しようとする課題】本発明の目的は、音声アプリケーションの設計および実装が、IPテレフォニー・システムの設計および実装から分離した状態を維持するVoIPベースの音声システムを提供することである。本発明のさらなる目的は、IPテレフォニー・システムから、VoIPネットワークを介して音声入力を受信可能なVoIP対応音声サーバを提供することである。また、本発明の目的は、VoIPネットワーク中のテレフォニー・ゲートウェイ・サーバに音声アプリケーションを結合する方法を提供することである。また、本発明の目的は、VoIPベースの音声システムと、VoIP対応音声サーバと、VoIPネットワーク、テレフォニー・ゲートウェイ・サーバ、および音声アプリケーションへの標準準拠インタフェースを使用してテレフォニー・ゲートウェイ・サーバに音声アプリケーションを結合する方法とのそれぞれを提供することである。

【0009】

【課題を解決するための手段】本発明の以上の目的およびその他の目的は、VoIPテレフォニー・ゲートウェイ・サーバと、それぞれが音声アプリケーションを含む少なくとも1つの音声サーバと、VoIPテレフォニー・ゲートウェイ・サーバと音声サーバとの間のVoIP対応呼制御インタフェースと、VoIPテレフォニー・ゲートウェイ・サーバと少なくとも1つの音声サーバ中の音声アプリケーションとの間のVoIP通信パスとを含むVoIPベースの音声システムにおいて、達成される。VoIPベースの音声システムにおいて、VoIPテレフォニー・ゲートウェイ・サーバと音声アプリケーションとは、VoIP対応呼制御インタフェースを介してVoIP通信パスを構築することが可能である。

【0010】動作時において、VoIPテレフォニー・ゲートウェイ・サーバは、テレフォニー・インタフェースから音声信号を受信し、音声信号をデジタル化音声データにデジタル化し、デジタル化音声データをV

oIP対応パケットに圧縮し、VoIP対応パケットを、少なくとも1つの音声サーバ中の音声アプリケーションへ、VoIP通信パスを介して、VoIPプロトコルを使用して送信することが可能である。それに対応して、音声アプリケーションは、VoIP対応パケットを受信し、デジタル化音声データをVoIP対応パケットから再構成し、デジタル化音声データを音声テキスト変換することが可能である。さらに、音声アプリケーションは、テキストを音声合成してデジタル化音声データを生成し、デジタル化音声データをVoIP対応パケットにカプセル化し、VoIP対応パケットは、通信パスを介して、VoIPテレフォニー・ゲートウェイ・サーバへ送信することが可能である。その後、VoIPテレフォニー・ゲートウェイ・サーバは、VoIP対応パケットを受信し、デジタル化音声データを、VoIP対応パケットから再構成し、デジタル化音声データを、テレフォニー・インタフェースを介して送信することが可能である。

【0011】本発明の一態様において、VoIPテレフォニー・サーバは、テレフォニー・インタフェースとVoIPゲートキーパーとを含むことが可能である。VoIPゲートキーパーは、テレフォニー・インタフェースを介して音声呼を受信し、それに応答して、VoIPゲートキーパーは、複数の音声サーバの中から1つの音声サーバを選択することが可能である。音声サーバが選択されると、VoIPゲートキーパーは、選択された音声サーバ中のVoIP対応音声アプリケーションに対して、音声呼が受信された旨をアラートすることが可能である。

【0012】本発明の別の態様において、音声サーバは、音声認識エンジンと、テキスト音声変換エンジンと、VoIPテレフォニー・ゲートウェイ・サーバを介して音声呼接続を構築する呼制御インタフェースと、音声データ・パスとを含むことが可能である。なお、音声データ・パスは、音声データを、構築された音声呼接続を介して、音声認識エンジンにストリーミングすることが可能である。同様に、音声データ・パスは、テキスト音声変換エンジンから構築された音声呼接続を介して、音声データをストリーミングすることが可能である。

【0013】本発明の更に別の態様において、音声アプリケーションは、音声ブラウザでもよい。音声ブラウザは、VoIP通信パスを介して受信される音声コマンドに応答してWebコンテンツを検索することが可能である。また、音声ブラウザは、音声データは、検索されたWebコンテンツを音声合成して音声データを生成することが可能である。また、音声ブラウザは、音声データを、VoIP通信パスを介して、VoIPテレフォニー・ゲートウェイ・サーバへ送信することが可能である。Webコンテンツは、VoiceXMLドキュメントで

もよい。

【0014】音声サーバは、VoIPテレフォニー・ゲートウェイ・サーバ、VoIP通信パス、および音声アプリケーションへの標準ベースのインタフェースを使用して実装することが好ましい。具体的には、音声サーバは、音声認識エンジンと、テキスト音声変換エンジンと、JSAPI音声インタフェースと、JTAPIテレフォニー・インタフェースと、JMFメディア・インタフェースとを含むことが可能である。JTAPIテレフォニー・インタフェースは、デジタル音声データを、VoIPテレフォニー・ゲートウェイ・サーバと音声アプリケーションとの間でトランスポートするための音声呼接続を構築することが可能である。JMFメディア・インタフェースは、デジタル音声データを、音声アプリケーションと音声呼接続との間でトランスポートするためのデータ・パスを構築することが可能である。JSAPI音声インタフェースは、デジタル化音声データを、音声アプリケーションから音声認識エンジンに伝達することが可能である。同様に、JSAPI音声インタフェースは、音声合成化音声データを、テキスト音声変換エンジンから、音声アプリケーションへ伝達することが可能である。

【0015】また、本発明は、VoIP通信パスを介してVoIPテレフォニー・ゲートウェイ・サーバと通信するように構成することが可能な音声アプリケーションを含むことができるVoIP対応音声サーバ中で実現可能である。また、VoIP対応音声サーバは、VoIPテレフォニー・ゲートウェイ・サーバへのVoIP対応呼制御インタフェースを含み、VoIP対応呼制御インタフェースは前記VoIP通信パスを構築することが可能である。動作時において、VoIP対応パケットを、音声アプリケーションは、VoIPテレフォニー・ゲートウェイ・サーバから、VoIP通信パスを介して受信可能である。その後、デジタル化音声データは、VoIP対応パケットから再構成することが可能であり、デジタル化音声データは、音声-テキスト変換されることが可能である。さらに、テキストは、デジタル化音声データに合成することが可能であり、デジタル化音声データは、VoIP対応パケットにカプセル化することが可能であり、VoIP通信パスを介して、テレフォニー・ゲートウェイ・サーバに送信することが可能である。

【0016】VoIP対応音声サーバの別の態様において、VoIP対応音声サーバは、音声認識エンジンと、テキスト音声変換エンジンと、音声データ・パスとを含むことが可能である。音声データ・パスは、音声データを、構築された音声呼接続を介して、音声認識エンジンにストリーミングすることが可能である。また、音声データ・パスは、テキスト音声変換エンジンから、構築された音声呼接続を介して、音声データをストリーミ

ングすることが可能である。

【0017】音声アプリケーションは、音声ブラウザであることが好ましい。音声ブラウザは、VoIP通信パスを介して受信される音声コマンドに応答してWebコンテンツを検索することが可能である。また、音声ブラウザは、検索されたWebコンテンツを音声合成して音声データを生成することが可能である。その後、音声ブラウザは、音声データを、VoIP通信パスを介して、VoIPテレフォニー・ゲートウェイ・サーバへ送信することが可能である。Webコンテンツは、VoiceXMLドキュメントでもよい。

【0018】VoIP対応音声サーバは、VoIPテレフォニー・ゲートウェイ・サーバ、VoIP通信パス、および音声アプリケーションへの標準ベースのインタフェースを使用して実装することが好ましい。具体的には、VoIP音声サーバは、JTAPIテレフォニー・インタフェースを含み、デジタル音声データを、前記VoIPテレフォニー・ゲートウェイ・サーバと前記音声アプリケーションとの間でトランスポートするための音声呼接続を構築することが可能である。さらに、VoIP対応音声サーバは、JMFメディア・インタフェースを有して、デジタル音声データを、音声アプリケーションと音声呼接続との間でトランスポートするためのデータ・パスを構築することが可能である。また、VoIP対応音声サーバは、デジタル化音声データを音声アプリケーションから音声認識エンジンに伝送することと、音声合成化音声データをテキスト音声変換エンジンから音声アプリケーションに伝送することと、の両方を行うためのJSAPI音声インタフェースを有することが可能である。

【0019】また、本発明は、音声アプリケーションを、VoIPネットワーク中のテレフォニー・ゲートウェイ・サーバに結合する方法を含む。この方法は、VoIPテレフォニー・ゲートウェイ・サーバとのVoIP通信パスを構築し、音声アプリケーションを、構築されたVoIP通信パスを介してテレフォニー・ゲートウェイ・サーバと通信するように構成するステップを含むことが可能である。さらに、VoIP対応パケットは、構築されたVoIP通信パスを介して、テレフォニー・ゲートウェイ・サーバから受信することが可能である。デジタル化音声データは、VoIP対応パケットから再構成することが可能であり、その後デジタル化音声データは、音声-テキスト変換することが可能である。さらに、この方法は、テキストをデジタル化音声データに合成し、デジタル化音声データをおIP対応パケットにカプセル化し、音声データを、VoIP通信パスを介して、テレフォニー・ゲートウェイ・サーバへ送信するステップを含むことが可能である。

【0020】好適な実施形態において、本方法は、Webコンテンツを、VoIP通信パスを介して受信され音

声認識された音声コマンドに応答して検索し、検索されたWebコンテンツを音声データに合成し、音声データを、VoIP通信パスを介して、テレフォニー・ゲートウェイ・サーバへ送信する工程をさらに含むことが可能である。Webコンテンツは、VoiceXMLドキュメントでもよい。

【0021】

【発明の実施の形態】本発明は、VoIPベースの音声システムであり、音声サーバはVoIPネットワーク中のテレフォニー・ゲートウェイ・サーバに結合することが可能である。テレフォニー・ゲートウェイ・サーバは、たとえば公衆交換電話網(PSTN)やサービス総合デジタル網(ISDN)などの外部電話網からの呼を受信できる。音声サーバは、テレフォニー・ゲートウェイ・サーバ中の音声呼から発信されVoIP通信パスを介して入力される実時間音声を受信可能である音声アプリケーションを備えることが可能である。同様に、音声アプリケーションは、音声合成音声データを、VoIP通信パスを介して、テレフォニー・ゲートウェイ・サーバへ、そして最終的には、外部電話網中の最終点へ伝送できる。特に、この音声アプリケーションは、音声呼を介して音声ブラウザ・コマンドを受信でき、音声アプリケーションはそれに応答して、外部ウェブ・サーバからWebコンテンツを検索可能である。さらに、このWebコンテンツは音声合成され、音声呼の一部でもあるVoIP通信パスを介して伝送される。好適実施形態において、このWebコンテンツは音声XMLドキュメントでもよい。

【0022】図1は、好適実施形態によるVoIPベースの音声システムを図示するものである。当業界では周知のように、VoIP仕様は、たとえばH.323、SIP、MGCPなどのいくつかの公開されている標準を使用して実装することが可能である。ただし、本発明は、VoIPの特定の実装に関して限定されるものではないが、本発明はH.323を実装するものとする。図1に図示するように、動作において、ユーザは、電話装置1を使用して音声呼を開始できる。音声呼は、たとえばPSTNやISDNなどの電話網2を介してVoIPテレフォニー・ゲートウェイ・サーバ3との接続を試みることが可能である。VoIPテレフォニー・ゲートウェイ・サーバ3は、音声呼の意図する受信先のアドレスを、VoIPネットワーク4中に存在する装置(この例では、VoIP対応音声サーバ5)のIPアドレスに変換できる。

【0023】その後、VoIPテレフォニー・ゲートウェイ・サーバ3は、VoIP対応音声サーバ5が受け入れ可能な音声呼をVoIP対応音声サーバ5に通知する。音声呼の受信時、VoIP対応音声サーバ5は、VoIPテレフォニー・ゲートウェイ・サーバ3とVoIP対応音声サーバ5との間にVoIP通信パスを構築し

て、音声データのVoIP対応パッケージが、VoIPテレフォニー・ゲートウェイ・サーバ3とVoIP対応音声サーバ5との間でトランスポート可能とする。このようにして、電話装置1から発信された音声データは、VoIP対応音声サーバ5中で受信および処理される。同様に、VoIP対応音声サーバ5から発信された音声データを、電話装置1に返送することができる。

【0024】なお、好適実施形態において、VoIP対応音声サーバ5は、電話装置1から発信される音声コマンドを受信して、データ通信網6中のWebサーバ7からWebコンテンツを検索することが可能である。特に、Webコンテンツ8は、VoiceXMLドキュメント8でもよい。それに応答して、音声対応音声サーバ5は、Webサーバ7からVoiceXMLドキュメント8を検索でき、VoiceXMLドキュメント8中に含まれる命令にしたがって、音声データを合成できる。その後、合成された音声データは、VoIPネットワーク4を介して、VoIPテレフォニー・ゲートウェイ・サーバ3へ、そして最終的には電話装置1にトランスポートすることが可能である。

【0025】図2は、図1のVoIPネットワーク4を図示するもので、VoIPテレフォニー・ゲートウェイ・サーバ3のより詳細な図を含む。図2に示すように、VoIPテレフォニー・ゲートウェイ・サーバ3は、電話装置1から、電話網2を介してテレフォニー・インタフェース11へ音声呼を受信可能である。テレフォニー・インタフェース11は、音声呼の意図する受信先のアドレスのアドレス変換を行い、それにしたがってその音声呼を送信可能である。さらに詳細に記せば、好適実施形態において、VoIPゲートキーパー14は、VoIPテレフォニー・ゲートウェイ・サーバ3中に組み込まれており、VoIPテレフォニー・ゲートウェイ・サーバ3へ呼管理機能を提供する。特に、VoIPゲートキーパー14は、負荷平衡化を行い、VoIP対応音声サーバ5が確実に音声呼を受信可能なようにする。

【0026】したがって、テレフォニー・インタフェース11における音声呼の受信時、呼制御は、呼制御インタフェース13を介してVoIPゲートキーパー14へ渡される。なお、VoIPゲートキーパー14は、データバス17を介してVoIPテレフォニー・ゲートウェイ・サーバ3の他のコンポーネントと通信可能である。さらに、呼制御インタフェース13は、VoIPゲートキーパー14中に含まれ、VoIPゲートキーパー14を介して処理される音声呼の確立、進行、および終了を制御できる。VoIPの好適な実装は、RTPベースのH.323標準の実装であるため、呼制御インタフェース13は、H.323ベースの呼制御インタフェースである。

【0027】その後、制御は、VoIPゲートキーパー14に渡され、拡張呼管理機能15を使用する呼プロセ



ッサ16は、V○IPネットワーク4中の各V○IP対応音声サーバ5のステータスを調べ、当該音声呼を受信するのに最適なV○IP対応音声サーバ5を識別する。その結果として、V○IPゲートキーパーは、最適なV○IP対応音声サーバ5を選択でき、当該音声呼の被選択V○IP対応音声サーバ5に対してアラートを発生できる。

【0028】アラートを受信すると、選択されたV○IP対応音声サーバ5は、V○IPテレフォニー・ゲートウェイ・サーバ3とV○IP対応音声サーバ5との間にV○IP通信パスを構築でき、それを介してV○IP対応パケットを伝送することができる。その結果、テレフォニー・インタフェース11は、当該音声呼に含まれる音声信号をデジタル化音声データにデジタル化し、そのデジタル化音声データをV○IP対応パケットに圧縮し、そのV○IP対応パケットをV○IP対応音声サーバへ、V○IP通信パス18を介して、V○IPプロトコルを使用して送信できる。

【0029】なお、本発明は、V○IPテレフォニー・ゲートウェイ・サーバ3の特定の構成に限定されるものではない。特に、V○IPゲートキーパー14がV○IPテレフォニー・ゲートウェイ・サーバ3の残りのコンポーネントとは別のエンティティとして図示されているが、本発明をそれに限定することを意図するものではない。図2のV○IPゲートキーパー14の配置は、図示目的のみを意図するものである。さらに、V○IPゲートキーパーとの組み合わせられたV○IPテレフォニー・ゲートウェイ・サーバ3に関する本発明の範囲は、V○IPテレフォニー・ゲートウェイ・サーバ3が音声呼を受信可能であり、V○IPゲートキーパー14はV○IPネットワーク4中の当該音声呼に対する最適受信先を識別することによって、呼管理を行うことができる点にのみ限定されるものである。

【0030】図3は、図1のV○IP対応音声サーバ5の好適アーキテクチャを図示するものである。V○IP対応音声サーバ5は、中央演算ユニット(CPU)と、ランダム・アクセス・メモリ(RAM)21や、たとえばハードディスク・ドライブ(HDD)などの固定ストレージ22などの内部メモリ装置を備える従来のネットワーク・サーバ中に実装することが可能である。V○IP対応音声サーバ5は音声対応であるため、音声回路(図示せず)も含み、V○IP対応音声サーバ5に音声処理機能を提供できる。

【0031】V○IP対応音声サーバ5は、固定ストレージ22中に、様々なアプリケーション・プログラムを実行させるためのオペレーティング・システム23を格納可能である。さらに、固定ストレージ22は、そこに、音声アプリケーション24とV○IPテレフォニー・モジュール25とを格納することが可能である。オペレーティング・システム23は、たとえばMicros

oft Windows(R) NT、Sun SolarisまたはDebian Linuxなどの最適なオペレーティング・システムを備えることが可能である。なお、本発明は、オペレーティング・システム23に関係して、音声アプリケーション24とテレフォニー・モジュール25との構成に関して限定されるものではない。むしろ、それぞれは、様々な組み合わせにおいて他のものと統合することが可能である。たとえば、V○IPテレフォニー・モジュール25は、オペレーティング・システム23中で統合することが可能である。あるいは、V○IPテレフォニー・モジュール25は、オペレーティング・システム23から独立した状態を維持できる。

【0032】また、本発明は、V○IPテレフォニー・モジュール25、音声アプリケーション24、およびそれらのコンポーネントの格納場所に限定されるものではない。むしろ、本発明は、より複雑な分散システム中に実装することができ、そのシステム中において、様々なコンポーネントが複数のネットワーク・サーバ中に常駐し、互いにリモートなプロセス・アドレス空間で実行し、各アプリケーションは、たとえば、TCP/IPなどの周知のプロセス間通信機構によって、他のアプリケーションと通信する。V○IP対応音声サーバ5のブートストラップ時、オペレーティング・システム23はRAM21にロード可能である。その後、音声アプリケーション24とV○IPテレフォニー・モジュール25との両方は、RAM21にロードし、実行できる。実行されると、V○IP対応音声サーバ5は、V○IP通信パスを介して音声呼およびその後のデータを受信するように構成される。

【0033】音声アプリケーション24は、音声認識エンジン34とテキスト-音声変換エンジン35とを含む。動作時において、V○IP対応音声サーバ5は、V○IP対応パケットを受信し、そのV○IP対応パケットからデジタル化音声データを再構成し、音声認識エンジン34中でデジタル化音声データを音声-テキスト変換する。一方、音声アプリケーション24は、テキスト-音声変換エンジン35において、テキストをデジタル化音声データに合成し、デジタル化音声データをV○IP対応パケットにカプセル化し、V○IP対応パケットを、V○IP通信パス18を介してV○IPテレフォニー・ゲートウェイ・サーバ3に送信することが可能である。

【0034】好適実施形態において、音声アプリケーションは、音声ブラウザ30を含む。なお、音声ブラウザ30は、V○IP通信パス18を介して受信される音声コマンドに回答してWebコンテンツを取り出し、音声認識エンジン34によって、音声-テキスト変換が行われ、音声ブラウザ30により解釈することが可能である。また、音声ブラウザ30は、受信されたWebコン



テンツをテキスト音声変換エンジン35に送信し、音声合成を行った後、音声合成化音声データを、V o I P通信バス18を介して、V o I Pテレフォニー・ゲートウェイ・サーバ3に送信することが可能である。特に、WebコンテンツはV o i c e XMLドキュメント8でもよい。

【0035】音声アプリケーションは、V o I P通信と音声認識および音声合成機能への標準準拠のインタフェースを使用して実装することが好ましい。さらに詳細に記せば、音声アプリケーション24は、音声認識エンジン34およびテキスト音声変換エンジン35と音声ブラウザ30との間のJ S A P I 音声インタフェース33を含むことが可能である。また、音声アプリケーション24は、テレフォニー・モジュール25と音声ブラウザ30との間のJ T A P I テレフォニー・インタフェース31を含むことが可能である。また、音声アプリケーション24は、テレフォニー・モジュール25と音声ブラウザ30との間にJ M F メディア・インタフェース32を含むことが可能である。

【0036】J T A P I テレフォニー・インタフェース31は、音声ブラウザ30が使用して、デジタル音声データを含むV o I P対応パケットをトランスポートするために、テレフォニー・ゲートウェイ・サーバ3と音声アプリケーション24との間に音声呼接続を構築することが可能である。J M F メディア・インタフェース32は、デジタル音声データを含むV o I P対応パケットをトランスポートするために、音声アプリケーション24と音声呼接続との間にV o I P通信データバスを構築することが可能である。J S A P I 音声インタフェース33は、デジタル化音声データを、音声アプリケーション24から音声認識エンジン34へ伝達することが可能である。同様に、J S A P I 音声インタフェース33は、音声合成化音声データを、テキスト音声変換エンジン35から音声アプリケーション24へ伝達することが可能である。

【0037】本発明のV o I Pベースの音声システムによって、ユーザは、電話を使用してWebサイトにアクセス可能となる。ユーザは、WebサイトとWebサイトの特徴を説明するプロンプトを音声で提示される。したがって、V o I Pベースの音声システムの表現は、統合音声応答システムと同様である。Webサイトの表示中、ユーザは、V o I Pベースの音声システムに口頭でコマンドを供給して、オプションの選択および情報の入力を行って、Webベースのフォームを完成させることが可能である。有効なことに、V o I Pベースの音声システムは、ページ記述言語として、V o i c e XMLを有するWebコンテンツを検索可能である。

【0038】本発明の特別な要素は、音声サーバであり、この音声サーバは、周知の標準V o I PプロトコルであるH. 323を使用して音声メッセージおよび制御

メッセージを送受信できる。V o I Pプロトコルの使用によって、音声サーバは、音声システムの他の要素から分離した状態を維持でき、したがって音声サーバは、より良好に最適化することが可能となる。また、分離された音声サーバを使用すると、音声サーバは、複数の使用可能なテレフォニー・ハードウェアの1つとの対話の詳細から解放されるため、開発工程が非常に簡潔となる。最後に、分離されたV o I P対応音声サーバの使用によって、音声サーバは、特別な開発がなくても、標準H. 323プロトコル、またはその他のV o I PプロトコルをサポートするV o I Pテレフォニー・システムとともに使用することが可能である。

【0039】新規の構成にしたがってV o I Pベースの音声システムには、音声サーバを、電話網とのインタフェース接続を行うテレフォニー・ゲートウェイ・サーバ中のテレフォニーのハードウェアおよびソフトウェアの詳細から分離するという利点がある。さらに、本発明の好適実施形態は、標準プロトコルを導入して、音声および呼制御情報を音声サーバにトランスポートし、それによって、その設計を単純化および最適化する。また、音声サーバへ音声および呼制御情報をトランスポートするために標準プロトコルを使用すると、音声サーバは、標準V o I Pプロトコルをサポートする様々なテレフォニー・ハードウェアおよびソフトウェアとともに動作することが可能となる。

【0040】本発明は、音声アプリケーションをV o I Pテレフォニー・ゲートウェイ・サーバに結合する方法を含むことが可能である。本発明は、ハードウェア、ソフトウェア、またはハードウェアとソフトウェアとの組み合わせにおいて、実現することが可能である。本発明による機械可読ストレージは、1台のコンピュータ・システム中で一括して、または異なる要素をいくつかの相互接続されたコンピュータ・システム間に分散して実現することが可能である。本願明細書において、説明される方法を実施するために適応されたあらゆる種類のコンピュータ・システムまたは他の装置は許容可能である。

【0041】ハードウェアとソフトウェアの典型的な組み合わせは、ロードされ実行されたときに、コンピュータ・システムを制御して本願明細書で説明された方法を実施するようにするコンピュータ・プログラムを有する汎用コンピュータ・システムである。また、本発明は、本願明細書中で説明される方法の実施を可能とする全ての機能を含み、コンピュータ・システム中にロードされるとそれらの方法を実行可能であるコンピュータ・プログラム製品中に組み込むことが可能である。

【0042】ここで言うコンピュータ・プログラムは、情報処理機能を有するシステムが特定の機能を直接、または以下的一方または両方の後に実行するようにすることを意図した1組の命令のあらゆる言語、コードまたは表記におけるあらゆる表現を意味する：(a) 別の言

語、コードまたは表記への変換、および(b)異なる形態における再生。本願明細書において開示される本発明は、上述のオペレーティング・システム23用の市販の開発ツールを使用してプログラマーによって実装され得るコンピュータ・プログラム中に組み込み可能である。

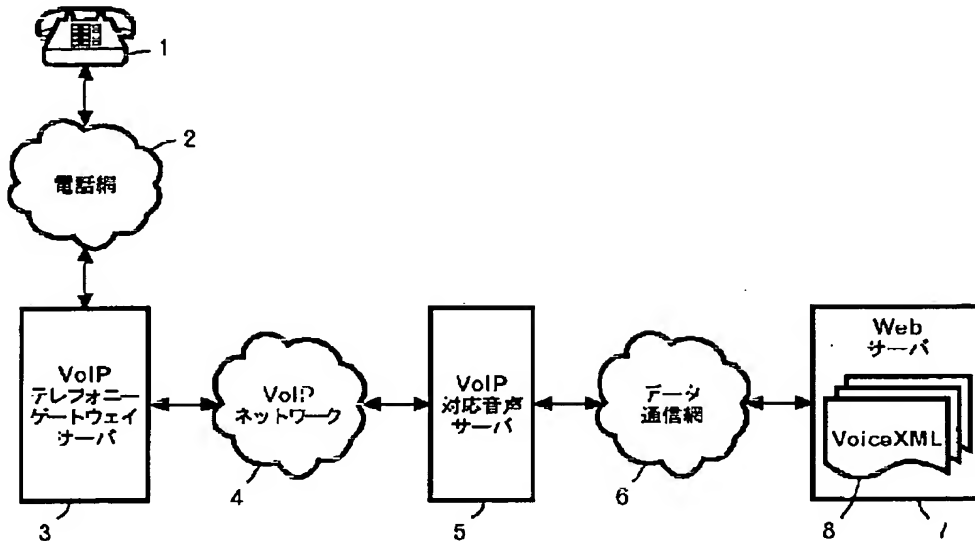
【図面の簡単な説明】

【図1】本発明によるVoIPベースの音声システムの概略図である。

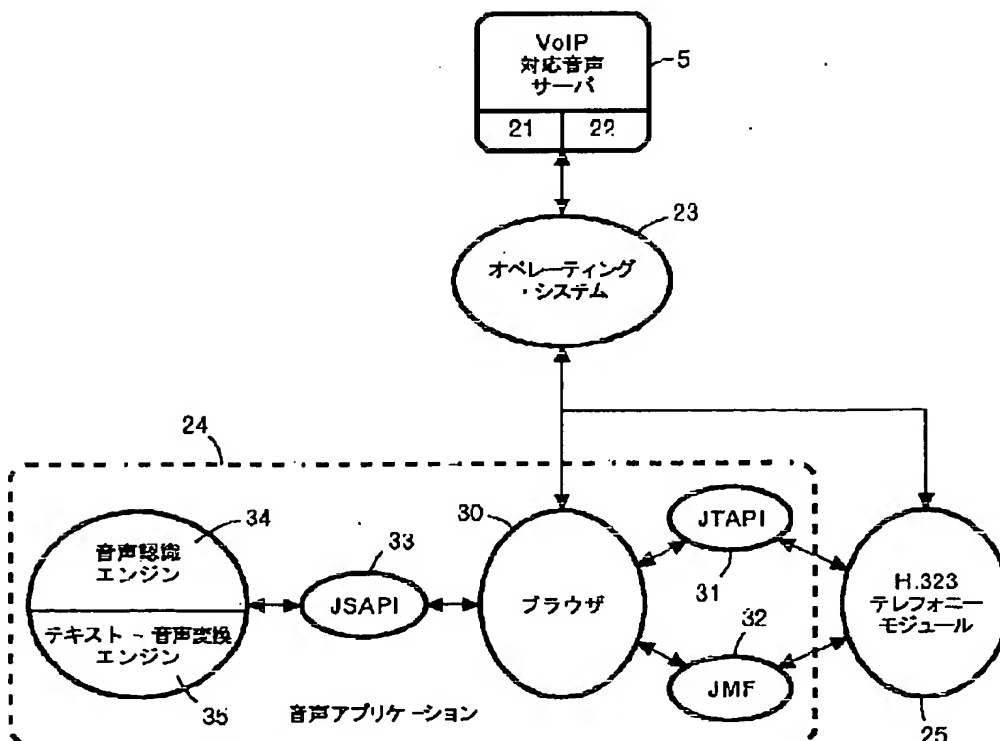
【図2】図1のVoIPテレフォニー・ゲートウェイ・サーバ用の好適アーキテクチャの図である。

【図3】図1の音声サーバ用の好適アーキテクチャの図である。

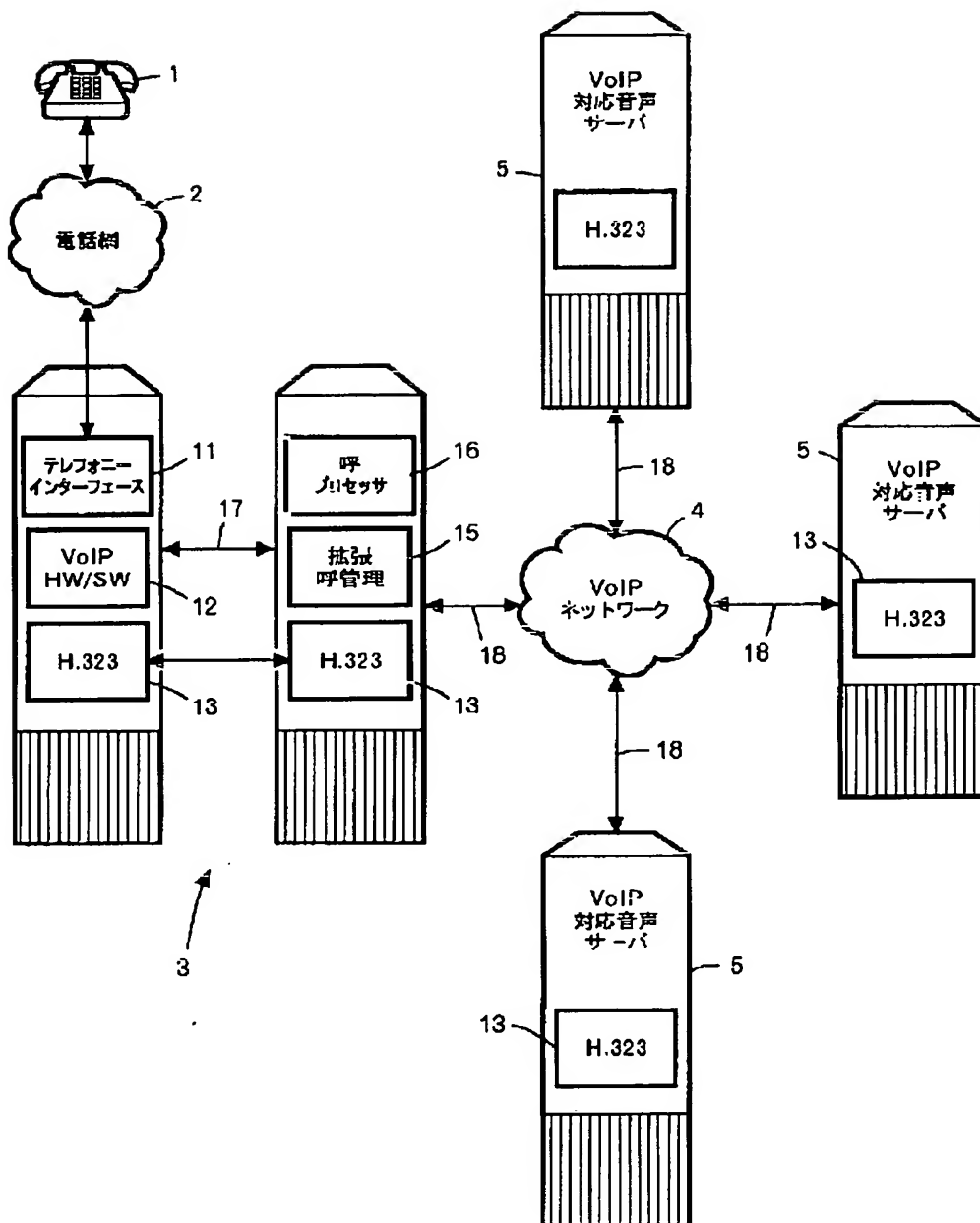
【図1】



【図3】



【図2】



フロントページの続き

(72)発明者 アン・エム・アルドス  
アメリカ合衆国33331、フロリダ州デイベー、  
ファルコンステージ・アベニュー 6380

(72)発明者 ジョセフ・セリ、ジュニア  
アメリカ合衆国33433、フロリダ州ボカ  
ラトン、ピア・ドロ 8492

(72)発明者 ブレット・ガバニ  
アメリカ合衆国33322、フロリダ州サンライズ、  
NW99番テラス 2932

(72)発明者 キリアコス・レオンチアズ  
アメリカ合衆国33431、フロリダ州ボカ  
ラトン、NW22 ストリート 353

(72)発明者 ブルース・ディ・ルーカス  
アメリカ合衆国10598、ニューヨーク州ヨ  
ークタウン・ハイツ、ミル・ボンド・ロー  
ド 2408

(72)発明者 デビッド・イー・レイチ  
アメリカ合衆国33458、フロリダ州ジュビ  
ター、インランド・コート 6684  
Fターム(参考) 5K030 HD03 KA19  
5K101 KK05 LL00 LL01 LL05 MM05  
NN08 NN16 RR05 SS08 UU19